Web Design Galleries: Please Give Me Similar Styles! A Claim for Ground Truth Datasets

Dimitri Masson[a], Alexandre Demeure[a], Zeina Abu-Aisheh[b], Gaëlle Calvary[b], Gilles Bisson[b]
[a] INRIA, Université Grenoble Alpes, LIG, F-38000 Grenoble, France
[b] Université Grenoble Alpes, LIG, F-38000 Grenoble, France

ABSTRACT
Web design galleries are extremely popular for searching inspiration in web design, but there is a lack of rich search functions. Recent works in the field have focused on style similarity browsing, where one hops from design to design based on their style similarity. In this paper, we claim for a study of the multiple dimensions of this notion of style, and of its perception by humans. We advocate for ground truth datasets based on a first experiment.

Author Keywords
Web gallery; layout; design; inspiration.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Design;

INTRODUCTION
Browsing examples on the web is a common design practice [1, 2]. It allows designers to follow trends and technological advances, and to reuse existing design pieces. This is particularly true in web design due to 1) its rapid technological and style evolution, 2) the small set of standard elements to be composed (*e.g.* header, navigation bar, content area, etc.), and 3) the possible reusability of design pieces thanks to the open source approach. Altogether these foster learning and building design through examples [3].

In practice, for scalability, web browsing is supported by search engines. They are performant for searching specific content-based elements but do not support browsing for design inspiration as they fail for instance to express background color or layout related (e.g., three columns) queries. Web design galleries are the common alternative.

Web design galleries are curated collections of websites. In general, they look like a design catalog where one can just flip through the pages. Despite being key for making from interactive galleries actual creative support tools [8], current galleries rarely provide rich search and multi-facets browsing functions (e.g., by categories, colors).

Inspired by advances in image retrieval, recent works in interactive design galleries have focused on similarity browsing (*i.e.,* the gallery is tailored to display elements that are similar to one or more elements that have been selected by the user). As such, these researches make the implicit hypotheses that 1) a perceptual similarity of styles might exist between designs, and 2) algorithms can compute such similarity. However, there is still no ground truth data on perceptual similarity in web design, and the similarity measures have never been assessed by designers.

Perceptual similarity is under explored. The notion of style needs to be investigated in order to identify its dimensions and their relative predominance. We argue that setting up ground truth datasets is essential and non-trivial .

THE NOTION OF STYLE IN THE LITERATURE
Style is quite ill-defined: "a distinctive quality, form, or type of something"[1]. The following subsections survey the definitions of web design style provided in design blogs, design galleries and academic interactive galleries.

From design blogs
Most design blogs report current styles and trends in web design as perceived by their contributors. For instance, according to inspectelement.com[2] and onextrapixel.com[3], in 2010 web design styles can be grouped according to the graphical elements they use: the type of images (illustrations and cartoons, large images and photo backgrounds, or photo-realism) or of textures or palette (patterns, two-tone color, or printed paper) for example. Finally some styles fall into more conceptual categories such as nature, abstract, retro or vintage look, or minimalism. In 2012, according to smashingmagazine.com[4]

NordiCHI '14, October 26 - 30 2014, Helsinki, Finland
Copyright 2014 ACM 978-1-4503-2542-4/14/10…$15.00.
http://dx.doi.org/10.1145/2639189.2641200

[1] http://www.merriam-webster.com/dictionary/style
[2] http://inspectelement.com/articles/10-fantastic-and-
[2] http://inspectelement.com/articles/10-fantastic-and-creative-web-design-styles/
[3] http://www.onextrapixel.com/2010/08/31/a-detailed-look-into-popular-styles-in-web-design/
[4] http://www.smashingmagazine.com/2010/05/04/web-

and webdesignledger.com[5], style trends currently puts the focus on layout and interaction, and seem to be more related to technological advances in HTML and CSS elements such as interaction techniques with responsive web design, fixed-position navigation, and multi-column menus; new possible shapes and effects such as circles, forked ribbons and banner graphics, zigzag borders, stitching and jQuery/CSS3/HTML5 animations; typography with letterpress, custom font faces, justified centered typography; finally type of images with 19th-century illustrations, big vector art and skeuomorphic features ("ornamentation or design features on an object that are copied from the object's form in another medium").

From Web galleries

A preliminary study of 20 "organized" web galleries shows that the key dimensions are category (13/20), color (8/20), date (6/20), and rough layout (4/20). As an example, *screenfluent.com* offers one of the largest variety of exploration means with a palette of 11 colors, 6 styles (classic, grunge, illustrated, minimalist, modern, retro) and 5 layouts (1 to 3 columns, mixed and fluid). *Unmatchedstyle.com* supplies an even richer exploration, but less structured, through the extensive use of tags to provide style support for tendency (e.g., clean, crisp, minimal), layout (e.g., grid, column, liquid, fixed navigation), 216 colors, technology (HTML 5, 960 Grid System) and interaction (e.g., parallax, tabbed menu). It even supports a preview through videos of the interaction with some of the websites.

Most web galleries use a mix of categories based on 1) the content or the purpose of the design and 2) its style. For instance, categories proposed by http://speckyboy.com/ does cover 1) Blogs, Ecommerce, Personal Portfolios, Web Design Agencies & Companies, Magazine & Newspaper, Web Application, Creative 404 designs; and 2) Dark, Minimal, Illustrative, Vintage & Retro, Typography, Large Background, Bright & Colorful.

Generally speaking, categories of the 20 studied webgalleries are mostly related to the type of the web site (e.g., portfolio, blog, app, splash), sometimes to its purpose (music, pictures, fashion), technology (CSS, HTML, Flash), origin (country, designer, company) or style (e.g., dark, retro, grunge).

Academic Interactive Galleries

Several techniques for browsing design galleries have been explored. Faceted Metadata Browsing [9] appears as very efficient for large galleries. However it requires the extraction of features to be understood and labeled by designers. An alternative is to browse by similarity.

Lee *et al.* [3] show that browsing galleries by similarity is more efficient than randomly. Interactive Design Galleries can roughly be divided into two groups: on one hand, the approaches where the generation of the design examples is controlled by the system; on the other hand, the approaches that compute similarity in between existing examples.

Generation of design examples. In Design Gallery [4] computer graphics are automatically produced and organized by varying an input vector. In Magellan [5], examples are produced by recombination and variation of websites using a genetic algorithm. Magellan makes evolve 3 aspects of style: the user interface elements, the background and foreground colors and textures, the size and location of the elements.

Similarity of existing examples. There are galleries of existing examples where similarity is deduced from a set of extracted features from the designs either manually as in Adaptive Ideas [3] or automatically as in d.tour [6]. In Adaptive Ideas, designers browse a set of examples through two variants of faceted metadata: either the gallery is composed of elements that are deemed similar to a specified element, or the gallery is composed of elements that maximize variety along one dimension of the style. The style is determined by the metadata manually assigned to each website, and is limited to background color, primary font, number of columns and visual density. In d.tour the style is automatically computed from the HTML DOM tree of the website. It is defined by 4 major aspects: *Page complexity* computed on the number of elements and their layout in the DOM tree, *visual density* based on Image and Text density, *Color* as expressed in Temperature (hue), Lightness and Saturation, and finally *Typography* (limited to font size). Although more sophisticated than in Adaptive Ideas, this definition of style still fails in several essential aspects, such as the layout (e.g., the number of columns, weight distribution), dynamic navigation or aesthetic elements (e.g., stitching, ribbon). Last but not least, if multiple elements are selected, their features are aggregated into an average function. As a result, selecting two designs such as one with a white background and another one with a black background results in displaying grey websites.

This overview of the state of the art shows that the notion of style is (1) multidimensional, (2) still vague, and (3) always addressed from a system point of view in the sense that the focus is set on what can be computed instead of what is really perceived. In the next section, we present a first experiment that aims at understanding what "being similar" means for humans. This step is mandatory for then developing the right algorithms to be used in webgalleries.

WHAT DOES "SIMILAR STYLES" MEAN FOR HUMANS?

The evaluation of style similarity of websites is not trivial. Indeed, as exposed in the previous section, website style relies on a complex combination of dimensions (layout, colors, etc.), making it difficult to clearly identify and isolate each of them. Moreover, due to its fuzzy nature, some aspect of style similarity may not end in a **consensus**

[5]http://webdesignledger.com/tips/web-design-trends-in-2012

between people (e.g. dominant color or layout similarity). In addition, people might not be **consistent** over time.

We present a first experiment to evaluate perceived layout similarity of websites. The focus on layout was done because it is a fundamental aspect of website design that is currently poorly supported by design galleries. The general principle of our experiment consists in 1) setting up a collection of web designs that illustrates the diversity of styles (this task could be alleviated by using tools such as [7] that automatically harvest and store Web Designs), and 2) exposing users to subsets of this collection and asking them to classify designs with respect to this style dimension. We now present in details the protocol we applied and discuss the results we obtained.

Protocol

The experiment used a collection of 211 screenshots from the frontpage of websites referenced by existing web galleries. The screenshot where automatically collected by a web crawler. The complete list of websites and the screenshots can be found at http://iihm.imag.fr/demeure/Dimitri/. We set up a web survey filled in by 164 participants (89 males, 46 females, 29 unknown) recruited by mailing lists.

Participants were first asked to complete a form to get their name, age, sex, job and expertise in web design (Likert scale from 1 to 5). They were then explained that the experiment was about evaluating websites similarity with regard to their layout. We proposed a short explanation presenting the notion of layout as *"the graphical structure of the page as well as the spatial organization of the areas"*. We explicitly asked participants to be careful that *"neither the color nor the web content should influence [their] decision"*.

Before starting the main session, participants were presented with the User Interface (UI) used for the experimentation (Figure 1). Websites were presented by groups of three (triplet in the rest of the paper). Buttons enabled participants to choose the two most similar designs among the three in terms of layout, but also to skip if they were unable to choose. They were also allowed to go back to the previous triplet in case they changed their mind or made an error.

The experiment consisted in exposing participants to 50 triplets. The 5 first triplets (reference triplets) were identical for every participant and were carefully selected to train participants and assess consensus between participants. The remaining triplets were chosen so that each triplet was evaluated by several participants to evaluate the **consensus**. The order of the screenshots in a triplet were balanced across the repetition. Moreover, 5 of the triplets were presented twice to the same participant to check the **consistency** of answers of the participant. Each couple of similar triplets was presented at a distance of at least 10 other triplets. Every 5 evaluations, participants were asked to explain the reasons why they made their choice.

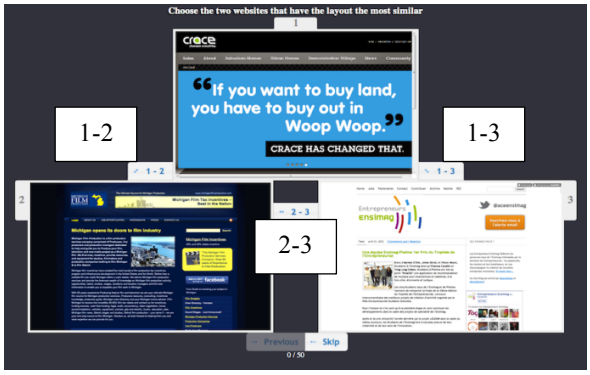

Figure 1. Test User Interface. "1-2", "1-3" and "2-3" enable participant to select most similar layouts.

Results

We analyzed the 693 triplets that were evaluated at least 7 times (Figure 223). Among them, 216 (~31%) triplets exhibit a consensus of at least 5/7. This means that there is a significant amount of consensual evaluation, which asses that perceptual layout similarity does exist between web designs. However, it is not possible to distinguish whether consensus emerged because websites were perceived as globally similar or because significant elements were perceived as similar.

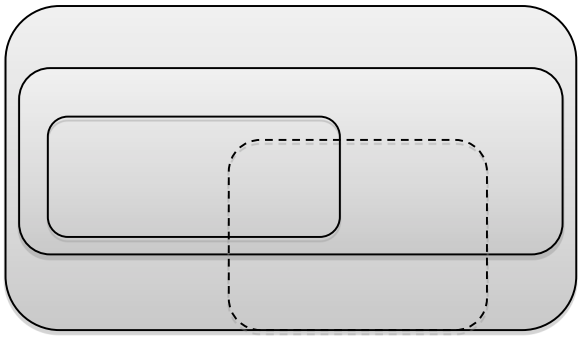

Figure 2. Summary of the triplets evaluations.

Participants skipped 1200 of the 8200 evaluations they did (164x50), however, only 5 of the 216 consensual triplets got a consensus on skipping the triplet. This indicates that most of the time, participants were able to find layout similarities even though it was not consensual (~69% of triplets evaluated at least 7 times). A possible explanation is that participants focused on different elements of websites.

Concerning the 5 reference triplets, 3 exhibit a consensus of at least 71% (82%, 76% and 75%), one of the others has a consensus of 64% and the last one has no consensus. As we carefully selected those 5 triplets, we were quite surprised that the consensus was not higher. On the other hand, this comforts us in the idea that evaluating layout similarity is not that trivial.

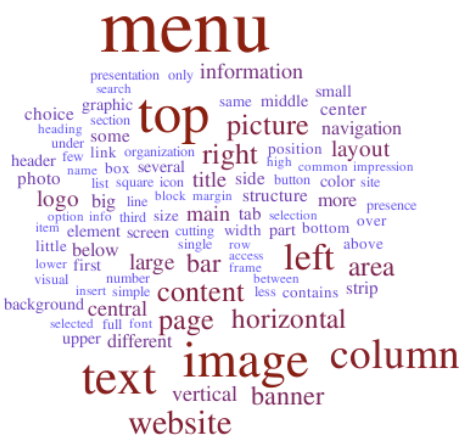

Figure 3. Word cloud generated from participants justifications of their choices.

We observed inconsistencies in participant answers when they were confronted twice to the same triplet but permuted differently. This was very common, 152 of the 164 users (~93%) made at least one inconsistency. On a total of 1102 triplets, 630 were tested for inconsistency with at least one user. This resulted in 341 triplets exhibiting at least one inconsistency (~54%). With respect to the 693 triplets having at least 7 evaluations, 465 were tested for inconsistency. This resulted in 253 triplets exhibiting at least one inconsistency (~54%). This may indicate that participants did change their mind along the test but more probably participants were not really sure about their choice. Indeed, among the 143 triplets having a high consensus and tested for consistency, only 53 were tested inconsistent (~37%) compared to ~54% for all the triplets tested at least 7 times.

Last, Figure 334 illustrates in a word cloud the justification of participants' choices. It highlights the importance of the **position** of the element (top, left, right), but also the importance of the **type** of the elements such as menu, navigation, logo, images or text.

CONCLUSION AND DISCUSSION

Web design galleries are a key element for inspiration and example retrieval. To improve on existing galleries current researches have focused on browsing by similarity. Those works rely on the implicit hypothesis there exists a perceptual similarity of style between designs and if so that it is possible to compute it. We argued that there is a need to construct a ground truth in terms of perceptual similarity to evaluate current research in web galleries. We presented an approach to build such a ground truth and applied it to one of the aspects style that is the layout.

This preliminary work shows that consensus on similar design between evaluators is possible, opening the possibility to find a perceptual similarity. **C**onsistency in participant answers is low which tends to indicate the difficulty to evaluate this similarity. We argue that it is necessary for the community to further explore this notion of perceptual similarity of style. In future work we intend to evaluate the salience of the different elements in the layout (e.g. logo position versus the number of columns used in the layout) and the relative importance of the style aspects (e.g. color scheme versus layout).

REFERENCES

1. Buxton, B. *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann.
2. Herring, S. R., Chang, C., Krantzler, J., and Bailey B. P. 2009. Getting inspired!: understanding how and why examples are used in creative design practice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '09).
3. Lee, B., Srivastava, S., Kumar, R., Brafman, R. and Klemmer, S.R. Designing with interactive example galleries. In *Proceedings of the 28th international conference on Human factors in computing systems*, 2010, pp. 2257-2266
4. Marks, J., Andalman, B., Beardsley, P. A., Freeman, W., Gibson, S., Hodgins, J., Kang, T., Mirtich, B., Pfister, H., Ruml, W., Ryall, K., Seims, J., Shieber, S. Design galleries: a general approach to setting parameters for computer graphics and animation. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, 1997, 389-400
5. Masson, D., Demeure, A., Calvary, G. Examples galleries generated by interactive genetic algorithms. In *Procedings of the Second Conference on Creativity and Innovation in Design.* ACM New York, NY, USA.
6. Ritchie, D., Kejriwal, A. A., Klemmer, S. R. d.tour: style-based exploration of design example galleries, In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, 2011.
7. Satyanarayan, A. and Lim, M. and Klemmer, S. A platform for large-scale machine learning on web design. In *Proceedings of the 2012 ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts*. 2012.
8. Shneiderman, B. Creativity support tools: accelerating discovery and innovation. In *Journal Communications of the ACM*, volume 50, 2007.
9. Yee, K.-P., Swearingen, K., Li, K., Hearst, M., Faceted metadata for image search and browsing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, April 05-10, 2003, Ft. Lauderdale, Florida, USA